



\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search. using Smith-Waterman algorithm  
Run on: Thu Feb 5 05:03:28 1998; MasPar time 1179.54 Seconds  
1181.343 Million cell updates/sec  
Tabular output not generated.  
  
Title: >US-08-770-564A-1  
Description: (1-981) from US08770564A.seq  
Perfect Score: 981  
N.A. Sequence: 1 CTGCAGAGGATAGAAAAAG.....ACTTAGTTCTGCTCTGCAG 981  
Comp: GACGTCTCCTATCTTTTC.....TGAATCAAGGACGAGACGTC  
  
Scoring table: TABLE default  
Gap 6  
  
Nmatch STD : Dbase 0; Query 0  
  
Searched: 430261 seqs, 710217276 bases x 2  
  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
  
Database: embl-new7  
1:BCT 2:FUN 3:GEN 4:HTG1 5:HTG2 6:HTG3 7:HTG4 8:HUM1  
9:HUM2 10:HUM3 11:INV1 12:INV2 13:ORG 14:MAM 15:VRT  
16:PLN 17:PRO1 18:PRO2 19:ROD 20:SYN 21:UNC 22:VIR  
Database: genbanka101  
23:BCT1 24:BCT2 25:BCT3 26:BCT4 27:BCT5 28:BCT6 29:BCT7  
30:BCT8 31:BCT9 32:BCT10 33:BCT11 34:BCT12 35:BCT13  
36:GEN1 37:GEN2 38:GEN3 39:GEN4 40:GEN5 41:GEN6 42:HTG1  
43:HTG2 44:HTG3 45:HTG4 46:HTG5 47:INV1 48:INV2 49:INV3  
50:INV4 51:INV5 52:INV6 53:INV7 54:INV8 55:INV9 56:INV10  
57:INV11 58:INV12 59:MAM1 60:MAM2 61:MAM3 62:VRT1  
63:VRT2 64:VRT3 65:VRT4 66:PAT1 67:PAT2 68:PAT3 69:PAT4  
70:PAT5 71:PAT6 72:PAT7 73:PHG 74:PLN1 75:PLN2 76:PLN3  
77:PLN4 78:PLN5 79:PLN6 80:PLN7 81:PLN8 82:PLN9 83:PLN10  
84:PLN11 85:PLN12 86:PRI1 87:PRI2 88:PRI3 89:PRI4  
90:PRI5 91:PRI6 92:PRI7 93:PRI8 94:PRI9 95:PRI10  
96:PRI11 97:PRI12 98:PRI13 99:PRI14 100:PRI15 101:PRI16  
102:PRI17 103:ROD1 104:ROD2 105:ROD3 106:ROD4 107:ROD5  
108:ROD6 109:ROD7 110:ROD8 111:ROD9 112:STR 113:SYN  
114:UNA  
Database: genbankb101  
115:VRL1 116:VRL2 117:VRL3 118:VRL4 119:VRL5 120:VRL6  
121:VRL7 122:VRL8 123:VRL9 124:VRL10 125:VRL11  
Database: genbank-new7  
126:BCT 127:GEN 128:HTG1 129:HTG2 130:INV 131:MAM  
132:VRT 133:PHG 134:PLN1 135:PLN2 136:PRI1 137:PRI2  
138:ROD 139:SYN 140:UNA 141:VRL

Database: u-emb151\_101  
142:part1 143:part2

Statistics: Mean 10.906; Variance 4.540; scale 2.402

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
	1	933	95.1	2420	72	I31750		Sequence 3 from patent	0.00e+00
	2	670	68.3	962	142	S79400		hTR=RNA component of	0.00e+00
	3	596	60.8	598	96	HSU85256		Human telomerase RNA,	0.00e+00
	4	514	52.4	554	72	I31748		Sequence 1 from patent	0.00e+00
	5	451	46.0	451	96	HSU86046		Human telomerase RNA	0.00e+00
	6	65	6.6	560	105	MMU33831		Mus musculus telomera	1.01e-32
	7	38	3.9	215	71	I28278		Sequence 5 from patent	2.28e-10
c	8	35	3.6	215	71	I28278		Sequence 5 from patent	3.99e-08
c	9	30	3.1	30	72	I31769		Sequence 22 from pate	1.45e-04
c	10	26	2.7	26	72	I31770		Sequence 23 from pate	6.35e-02
c	11	26	2.7	54404	5	HS366D1		Human DNA sequence **	6.35e-02
c	12	26	2.7	54404	128	HS366D1		Human DNA sequence **	6.35e-02
c	13	25	2.5	565	68	E04076		gDNA encoding envelop	2.68e-01
	14	25	2.5	1658	97	HUMA2PIG1		Human alpha-2-plasmin	2.68e-01
c	15	25	2.5	3510	32	PSESDSAB11		Pseudomonas sp. (stra	2.68e-01
c	16	25	2.5	3720	102	S57132		COL16A1-type XVI coll	2.68e-01
c	17	25	2.5	5387	97	HUMCOL16A		Homo sapiens alpha-1	2.68e-01
	18	24	2.4	201	66	A10158		S.griseus gene for pr	1.09e+00
	19	24	2.4	201	66	A10161		Synthetic DNA for pre	1.09e+00
c	20	24	2.4	201	66	A10162		Synthetic DNA for pre	1.09e+00
c	21	24	2.4	201	66	A10159		S.griseus gene for pr	1.09e+00
c	22	24	2.4	1157	122	HSEIR5A		Equine herpesvirus ty	1.09e+00
	23	24	2.4	1663	60	MVU92534		Mustela vison microsa	1.09e+00
	24	24	2.4	1877	108	MUSSEQUENC		Mus musculus HIC-5 mR	1.09e+00
c	25	24	2.4	4385	122	HSES1RO		Equine herpes virus t	1.09e+00
	26	24	2.4	4605	100	HUMPREP		Human gene for prepro	1.09e+00
	27	24	2.4	4606	102	S73906		adrenomedullin=potent	1.09e+00
c	28	24	2.4	4928	116	EHV1SGNS		Equine herpesvirus ty	1.09e+00
c	29	24	2.4	38793	18	MTCY270		Mycobacterium tubercu	1.09e+00
c	30	24	2.4	38793	31	MTCY270		Mycobacterium tubercu	1.09e+00
c	31	24	2.4	95257	45	HS999D10		Human DNA sequence **	1.09e+00
	32	24	2.4	139887	96	HUB384D8		Chromosome 22q13 BAC	1.09e+00
c	33	24	2.4	144406	44	HS104C13		Human DNA sequence **	1.09e+00
	34	24	2.4	150223	122	HSECOMGEN		Equine herpesvirus 1	1.09e+00
c	35	24	2.4	150223	122	HSECOMGEN		Equine herpesvirus 1	1.09e+00
	36	23	2.3	69	72	I41362		Sequence 142 from pat	4.24e+00
c	37	23	2.3	320	87	HS152D3R		H.sapiens CpG DNA, cl	4.24e+00
	38	23	2.3	565	68	E04076		gDNA encoding envelop	4.24e+00
	39	23	2.3	1595	104	MMNCAM1B		Mouse mRNA for neural	4.24e+00
c	40	23	2.3	1966	106	MUSCALCHAN		Mus musculus L-type c	4.24e+00
c	41	23	2.3	2500	76	D86598		Norway spruce mRNA fo	4.24e+00
	42	23	2.3	4552	19	MMNCAMR3		Mouse mRNA for 3'-end	4.24e+00
	43	23	2.3	4552	104	MMNCAMR3		Mouse mRNA for 3'-end	4.24e+00
	44	23	2.3	33769	5	HSAC2100		Genomic sequence from	4.24e+00
	45	23	2.3	33769	86	AC002100		Genomic sequence from	4.24e+00

#### ALIGNMENTS

RESULT 1  
LOCUS I31750 2420 bp DNA  
DEFINITION Sequence 3 from patent US 5583016.

ACCESSION I31750  
 NID g1822541  
 KEYWORDS .  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.  
 REFERENCE 1 (bases 1 to 2420)  
 AUTHORS Villeponteau,B., Feng,J., Funk,W. and Andrews,W.H.  
 TITLE Mammalian telomerase  
 JOURNAL Patent: US 5583016-A 3 10-DEC-1996;  
 FEATURES Location/Qualifiers  
 source 1..2420  
 /organism="unknown"  
 BASE COUNT 620 a 572 c 647 g 581 t  
 ORIGIN  
 Query Match 95.1%; Score 933; DB 72; Length 2420;  
 Best Local Similarity 99.1%; Pred. No. 0.00e+00;  
 Matches 972; Conservative 0; Mismatches 3; Indels 6; Gaps 5;  
 Db 1194 ctgcagaggatagaaaaaggccctgtataccctcaagtttagttcacctttaagaagg 1253  
 |||||||  
 Qy 1 CTGCAGAGGATAGAAAAAGWCCCTCTGATACCTCAAGTTAGTTCACCTTAAAGAAGG 60  
 Db 1254 tcggaagtaaagacgcaaagccttccggacgtgcggaggcaacgtccttcctcatg 1313  
 |||||||  
 Qy 61 TCGGAAGTAAAGACGCAAAGCCTTCCGGACGTGCGGAAGGGCACGTCCCTTCATG 120  
 Db 1314 gccggaaatgaaacttaattccgttcccccaaccagccgcggagagagtgactc 1373  
 |||||||  
 Qy 121 GCCGGAAATGAACTTAATTCCCGTTCCCCCAACCAGCCGCCGGAGAGAGTGAAC 180  
 Db 1374 tcacgagagccgcgagagtcaagttggcaatccgtgcggcgccgtcccttat 1433  
 |||||||  
 Qy 181 TCACGAGAGCCGCCGAGAGTCAGCTTGGCCAATCCGTGCGGTGGCCGCTCCCTTAT 240  
 Db 1434 aagccactcgcccgacgcgcaccgggtgcggagggtggcctgggggggtggcgc 1493  
 |||||||  
 Qy 241 AAGCCGACTCGCCCGCAGCGCACCGGGTTGCGGAGGGTGGCCTGGGAGGGGTGGTGGC 300  
 Db 1494 catttttgtaaccctaactgagaaggcgtaggcgcgtgcgtttgtcccccgcgc 1553  
 |||||||  
 Qy 301 CATTTTTGCTAACCTAACCTGAGAAGGGCGTAGGCGCCGTGCTTGTCCCCGCGCG 360  
 Db 1554 ctgttttctcgacttcaagcgccggaaaaggctcgccgtccgcaccgtt 1613  
 |||||||  
 Qy 361 CTGTTTTCTCGCTGACTTCAGCGGGCGAAAAGCCTCGGCCTGCCCTCCGGGACCGTT 420  
 Db 1614 cattcttagagcaaacaaaaatgtcagctgtggccgttgc--tcccggggacctgc 1671  
 |||||||  
 Qy 421 CATTCTAGAGCAAACAAAAATGTCAGCTGCTGGCCCGTTGCGCCCTCCGGGACCTGC 480  
 Db 1672 ggcgggtcgccgtcccgaccccgccgtggag-ccgcgggtcgccggggcttc 1730  
 |||||||  
 Qy 481 GGCGGGTGCGCTGCCAGCCCCGAACCCCGCTGGAGGCCGCGGTGGCCCGGGCTTC 540  
 Db 1731 tccggaggcaccactgccaccgcgaagagttggctgtcagccgcgggtctctcg 1790  
 |||||||  
 Qy 541 TCCGGAGGCACCCACTGCCACCGCGAAGAGAGTGGCTCTGTCAGCCGCGGGTCTCTCGGG 600  
 Db 1791 ggcgaggcgaggttac-cgtttcaggccgcaggaagaggaacggagcgagtcgc- 1847  
 |||||||  
 Qy 601 GGCAGGGCGAGGTTCAAGGCCCTTCAGGCCCGCAGGAAGAGGAACGGAGCGAGTCCCAGCG 660  
 Db 1848 cgccgcgcattccctgagctgtggacgtgcacccaggactcggtcacatgcagtt 1907  
 |||||||

Qy 661 CGCGGCGCGATTCCCTGAGCTGTGGGACGTGCACCCAGGACTCGGCTCACACATGCAGTT 720  
 Db 1908 cgcttcctgtggggggaaacccgatcgcatcgcatccgcacccctcgccggcagt 1967  
 |||||||  
 Qy 721 CGCTTCCCTGTTGGTGGGGGAACGCCGATCGTCGCATCGTCACCCCTGCCGGCAGT 780  
 |||||||  
 Db 1968 gggggcttgcaccccaaacactgactgactggccagtgctgcaattggcaggag 2027  
 |||||||  
 Qy 781 GGGGGCTTGTGAACCCCCAACCTGACTGACTGGCCAGTGTGCTGCAAATTGGCAGGAG 840  
 |||||||  
 Db 2028 acgtgaaggcacctccaaagtggccaaaatgaatggcagtgagccgggttgcctgga 2087  
 |||||||  
 Qy 841 ACGTGAAGGCACCTCCAAAGTCGGCCAAAATGAATGGCAGTGAGCCGGGTTGCCTGGA 900  
 |||||||  
 Db 2088 gccgttcctgcgtgggtctccgttttgcctttatgggttattac 2147  
 |||||||  
 Qy 901 GCCGTTCCCTGCGTGGGTTCTCCGCTTCCGCTTTGTTGCCTTTATGGTTGTATTAC 960  
 |||||||  
 Db 2148 aacttagttcctgctctgcag 2168  
 |||||||  
 Qy 961 AACTTAGTTCCCTGCTCTGCAG 981

RESULT 2  
 ID S79400 backbone; DNA; HUM; 962 BP.  
 AC S79400;  
 NI g1086943  
 DT 04-DEC-1995 (Rel. 45, Created)  
 DT 30-JAN-1997 (Rel. 50, Last updated, Version 3)  
 DE hTR=RNA component of telomerase [human, 293 cells, Genomic, 962  
 DE nt].  
 KW .  
 OS Homo sapiens (human)  
 OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 OC Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP 1-962  
 RX MEDLINE; 95381057.  
 RA Feng J., Funk W.D., Wang S.S., Weinrich S.L., Avilion A.A.,  
 RA Chiu C.P., Adams R.R., Chang E., Allsopp R.C., Yu J., Le S.,  
 RA West M.D., Harley C.B., Andrews W.H., Greider C.W.,  
 RA Villeponteau B.;  
 RT "The RNA component of human telomerase";  
 RL Science 269:1236-1241(1995).  
 CC NCBI gi: 1086943  
 FH Key Location/Qualifiers  
 FH  
 FT source 1..962  
 FT /organism="Homo sapiens"  
 FT /note="human"  
 FT misc\_RNA 1..962  
 FT /partial  
 FT /gene="hTR"  
 FT /note="telomerase RNA; Description: RNA component of  
 FT telomerase"  
 SQ Sequence 962 BP; 178 A; 257 C; 288 G; 239 T; 0 other;  
 Query Match 68.3%; Score 670; DB 142; Length 962;  
 Best Local Similarity 98.9%; Pred. No. 0.00e+00;  
 Matches 708; Conservative 0; Mismatches 2; Indels 6; Gaps 5;  
 Db 1 gggttgcggagggtggcctgggaggggtggtgccattttgcataaccctaactgag 60  
 |||||||  
 Qy 266 GGGTTGCGGAGGGTGGGCCTGGGAGGGGTGGTGCCATTGGTCAACCCTAACTGAG 325  
 |||||||  
 Db 61 aaggcgttaggcgcgtgtttgcctccgcgcgttttcgtacttcagcg 120

Qy	326	AAGGGCGTAGGCGCCGTCTTGCCTCCCCGCGCTGTTTCAGCTGACTTCAGCG	385
Db	121	ggcgaaaaagcctcgccctgcgccttccaccgttcatcttagagcaaaaaatgtc	180
Qy	386	GGCGGAAAGCCTCGGCCTGCCGCCTCCACCGTCATTCTAGAGCAAACAAAAATGTC	445
Db	181	agctgctggcccgttcgc--tccccgggacactgcggcggtgcgcctgcggccagccccga	238
Qy	446	AGCTGCTGGCCCGTTCGCCCTCCCAGGGACCTGCAGGGTCGCCTGCCAGCCCCGA	505
Db	239	accccgcctggag-ccgcggcgtggccgggctctccggaggcacccactgccaccgc	297
Qy	506	ACCCCGCCTGGAGGCCGGTCGGCCCGGGCTCTCCGGAGGCACCCACTGCCACCGCG	565
Db	298	aagagttggctctgtcagccgcgggtctctggggcgagggcgaggttcac-cgttac	356
Qy	566	AAGAGTTGGCTCTGTAGCCGCGGTCTCGGGGGCGAGGGCGAGGTTCAAGGCCTTTC	625
Db	357	aggccgcaggaagaggaacggagcgcgcgtccc-gc-cgcggcgcgattccctgagctgtgg	414
Qy	626	AGGCCGCAGGAAGAGGAACGGAGCGAGTCCCCGCGCGCGATTCCTGAGCTGTGG	685
Db	415	gacgtgcacccaggactcggtcacacatgcagttcgctttctgttggtggggaaacg	474
Qy	686	GACGTGCACCCAGGACTCGGCTCACACATGCAGTTCGCTTCCTGTTGGGGGGAACG	745
Db	475	ccgatcgtgcgcatccgtcacccctgcggcagtggtttgtgaaccccaaacctg	534
Qy	746	CCGATCGTGCATCCGTACCCCTGCCGGCAGTGGGGCTTGTGAACCCCAAACCTG	805
Db	535	actgactggccagtgtgctgcaaattggcaggagacgtgaaggcacccaaactgc	594
Qy	806	ACTGACTGGGCCAGTGTGCTGCAAATTGGCAGGAGACGTGAAGGCACCTCAAAGTCGGC	865
Db	595	caaatgaatggcagtgagccgggttgcctggagccgttctgcgtgggttctccgt	654
Qy	866	CAAAATGAATGGGCAGTGAGCCGGGTTGCCTGGAGGCCGTCCCTGCGTGGTTCTCCGT	925
Db	655	cttccgcctttgtgcctttatggtgtattacaacttagttcctgcgtcag	710
Qy	926	CTTCGCTTTGTTGCCTTATGGTTATTACAACCTAGTTCTGCTCTGCAG	981

RESULT 3

LOCUS HSU85256 598 bp DNA PRI 11-APR-1997

DEFINITION Human telomerase RNA, partial sequence.

ACCESSION U85256

NID g1932797

KEYWORDS .

SOURCE human.

ORGANISM Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;  
Homo.

REFERENCE 1 (bases 1 to 598)

AUTHORS Bryan,T.M., Marusic,L., Bacchetti,S., Namba,M. and Reddel,R.R.

TITLE The telomere lengthening mechanism in telomerase-negative immortal human cells does not involve the telomerase RNA subunit  
Hum. Mol. Genet. (1997) In press

REFERENCE 2 (bases 1 to 598)

AUTHORS Bryan,T.M. and Reddel,R.R.

TITLE Direct Submission

JOURNAL Submitted (13-JAN-1997) Children's Medical Research Institute, 214 Hawkesbury Rd., Westmead, NSW 2145, Australia

REFERENCE 3 (bases 1 to 598)

AUTHORS Bryan,T.M. and Reddel,R.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-APR-1997) Children's Medical Research Institute, 214  
 Hawkesbury Rd., Westmead, NSW 2145, Australia  
 REMARK Sequence update by submitter  
 FEATURES Location/Qualifiers  
 source 1..598  
     /organism="Homo sapiens"  
     /cell\_line="four immortal human cell lines (293, SUSM-1,  
     GM847, GM639)"  
     /chromosome="3q"  
 misc\_RNA 1..>598  
     /note="hTR; RNA component of telomerase; four sequence  
     differences found as compared to the sequence of earlier  
     submission of GenBank Accession Number S79400"  
     /product="telomerase RNA"  
 BASE COUNT 93 a 191 c 203 g 111 t  
 ORIGIN  
 Query Match 60.8%; Score 596; DB 96; Length 598;  
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;  
 Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Db 1 gggttgcggagggtggcctgggaggggtggtgccataaaaatgtctaaccctaacttag 60  
   |||||||  
 Qy 266 GGGTTGCGGAGGGTGGCCTGGGAGGGGTGGTGCCATTTTGTCTAACCTAACTGAG 325  
 Db 61 aaggcgtaggcgcgtgtttgtccccgcgcgttttctcgctgacttcacgc 120  
   |||||||  
 Qy 326 AAGGGCGTAGGCCGTGCTTTGCTCCCCGCGCGTGTGACTTCAGCG 385  
 Db 121 ggccggaaaagcctcgccgtccaccgttcatctagacaaacaaaaatgtc 180  
   |||||||  
 Qy 386 GGCGGAAAAGCCTCGGCCTGCCGCTTCCACCGTCATTCTAGAGCAAACAAAAATGTC 445  
 Db 181 agctgtggccgttcgccccccggggacctgcggcggtcgccctgcccagccccca 240  
   |||||||  
 Qy 446 AGCTGCTGGCCCGTTCGCCCTCCGGGACCTGCGGGTGCCTGCCAGCCCCGA 505  
 Db 241 accccgcctggaggccgcgtcgccccgggttctccggaggcacccactgccaccgc 300  
   |||||||  
 Qy 506 ACCCCGCCTGGAGGCCCGGTCGGCCCGGGCTCTCCGGAGGCACCCACTGCCACCGCG 565  
 Db 301 aagagttggctctgtcagccgcgggtctctggggcgagggcgaggttcaggcatttc 360  
   |||||||  
 Qy 566 AAGAGTTGGCTCTGTCAAGCCGGTCTCGGGGCGAGGGCGAGGTTCAAGCCCTTC 625  
 Db 361 aggccgcaggaaagaggaacggagcgagtcccgccggcgatccctgagctgtgg 420  
   |||||||  
 Qy 626 AGGCCGCAGGAAGAGGAACGGAGCGAGTCAGTCCCAGCGCGCGGCGATTCCCTGAGCTGTGG 685  
 Db 421 gacgtgcacccaggactcggtcacacatgcagttcgcttgcgtttgtggggaaacg 480  
   |||||||  
 Qy 686 GACGTGCACCCAGGACTCGGCTCACACATGCAGTTGCTTCTGTTGGTGGGGAAACG 745  
 Db 481 ccgatcgccatccgtcacccctgcggcaatggggcttgtgaaccccaaacctg 540  
   |||||||  
 Qy 746 CCGATCGTGCAGCATCCGTACACATGCAGTTGCTTCTGTTGGTGGGGAAACG 805  
 Db 541 actgactggccagtgtgctgcaaattggcaggagacgtgaaggcacctccaaagtgc 598  
   |||||||  
 Qy 806 ACTGACTGGGCCAGTGTGCTGCAAATTGGCAGGAGACGTGAAGGCACCTCCAAAGTCG 863

RESULT 4 I31748 554 bp DNA PAT 20-DEC-1996  
LOCUS

DEFINITION Sequence 1 from patent US 5583016.  
 ACCESSION I31748  
 NID g1822539  
 KEYWORDS .  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED Unclassified.  
 REFERENCE 1 (bases 1 to 554)  
 AUTHORS Villeponteau,B., Feng,J., Funk,W. and Andrews,W.H.  
 TITLE Mammalian telomerase  
 JOURNAL Patent: US 5583016-A 1 10-DEC-1996;  
 FEATURES Location/Qualifiers  
 source 1..554  
 /organism="unknown"  
 BASE COUNT 80 a 179 c 189 g 106 t  
 ORIGIN

Query Match 52.4%; Score 51.4; DB 72; Length 554;  
 Best Local Similarity 98.6%; Pred. No. 0.00e+00;  
 Matches 552; Conservative 0; Mismatches 2; Indels 6; Gaps 5;

Db	1 gggttgcggagggtgggcctgggaggggtggtgccattttgcataaccctaactgag	60
Qy	266 GGGTTGCGGAGGGTGGGCCTGGGAGGGGTGGTGCCATTGGTCAACCTAACTGAG	325
Db	61 aaggcgtaggcgcgtctttgcctccgcgcgtgttttcgtacttcacgcg	120
Qy	326 AAGGGCGTAGGCCTGCTTTGCTCCCCGCGCCTGTTTCGCTGACTTCAGCG	385
Db	121 ggcggaaaagcctcgccctgcgcgcgttcattctagagcaaaaaatgtc	180
Qy	386 GGCGGAAAAGCCTCGGCCTGCCGCCTCCACCCTGTTCTAGAGCAAACAAAAATGTC	445
Db	181 agctgctggcccgttcgc--tccccgggacctgcggcggtgcctgcccagccccga	238
Qy	446 AGCTGCTGGCCCCTCGCCCTCCGGGACCTGCAGGGCTGCCTGCCAGCCCCGA	505
Db	239 accccgcctggag-ccgcgtcgccggggcttcggagggcaccactgccaccgcg	297
Qy	506 ACCCCGCCTGGAGGCCGCGTCGGCCGGCTCTCCGGAGGCACCCACTGCCACCGCG	565
Db	298 aagagttggcttgtcagccgcgggtctcggggcgagggcgaggtcac-cgttc	356
Qy	566 AAGAGTTGGCTCTGTCAGCCGCAGGCTCTCGGGCGAGGGCGAGGTCAGGCCTTC	625
Db	357 aggccgcaggaagaggaacggagcgagtccc-gc-cgcggcgcatccctgagctgtgg	414
Qy	626 AGGCCGCAGGAAGAGGAACGGAGCGAGTCCCCGCGCGCGAGTCCCTGAGCTGTGG	685
Db	415 gacgtgcacccaggactcggtcacacatgcagttcgctttctgttggtggggaaacg	474
Qy	686 GACGTGCACCCAGGACTCGGCTCACACATGCAGTTGCTTCCGTGTTGGTGGGGGAACG	745
Db	475 ccgatcgtgcgcatccgtcacccctgcggcagtggtttgtgaaccccaaacctg	534
Qy	746 CCGATCGTGCATCCGTACCCCTGCCGGCAGTGGGGCTTGTGAACCCCCAACCTG	805
Db	535 actgactgggccagtgtgct 554	
Qy	806 ACTGACTGGGCCAGTGTGCT 825	

RESULT 5  
 LOCUS HSU86046 451 bp DNA PRI 27-JAN-1997  
 DEFINITION Human telomerase RNA (hTR) gene sequence.  
 ACCESSION U86046 S79400



Qy 686 GACGTGCACCCAGGACTCGGCTCACACATGC 716

RESULT 6  
LOCUS MMU33831 560 bp DNA ROD 28-SEP-1995  
DEFINITION Mus musculus telomerase RNA component gene.  
ACCESSION U33831  
NID g1000197  
KEYWORDS .  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;  
Murinae; Mus.  
REFERENCE 1 (bases 1 to 560)  
AUTHORS Blasco,M.A., Funk,W., Villeponteau,B. and Greider,C.W.  
TITLE Functional characterization and developmental regulation of mouse  
telomerase RNA  
JOURNAL Science 269 (5228), 1267-1270 (1995)  
MEDLINE 95381063  
REFERENCE 2 (bases 1 to 560)  
AUTHORS Blasco,M.A., Funk,W., Villeponteau,B. and Greider,C.W.  
TITLE Direct Submission  
JOURNAL Submitted (12-AUG-1995) Carol W. Greider, Cold Spring Harbor  
Laboratory, PO Box 100, 1 Bungtown Road, Cold spring Harbor, NY  
11724, USA  
FEATURES Location/Qualifiers  
source 1..560  
/organism="Mus musculus"  
/note="initially cloned from D3 embryonic stem cell  
library"  
snRNA 25..>453  
/note="The RNA is approximately 430 nucleotides long. The  
5' end has been determined, but 3' end has not been  
precisely mapped. Based on RT-PCR analysis and sequence  
conservation with the human homolog the 3' end is near  
position 453 in this sequence"  
/product="telomerase RNA component"  
misc\_feature 60..68  
/note="The template region of the RNA specifies the  
synthesis of the telomere sequence TTAGGG"  
BASE COUNT 90 a 183 c 188 g 99 t  
ORIGIN  
Query Match 6.6%; Score 65; DB 105; Length 560;  
Best Local Similarity 68.8%; Pred. No. 1.01e-32;  
Matches 256; Conservative 0; Mismatches 101; Indels 15; Gaps 12;  
Db 95 cttttgttctccgcccgttttctcgctacttcagcgggccaggaaagtccagac 154  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Qy 344 CTTTTGCTCCCCGCGCGTGTCTCGCTGACTTCAGCGGGCG-GAAAAGCCTCGGC 402  
Db 155 ctgcagcgggccacccggcgttcccagcct-caaaaacaacgtcagcgcaggagctcc 213  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Qy 403 CTGCCGCCTCCACC GTCAATTCTAGAGCAAACAAAAATGTCAGCTGCTGGCCCCTCG 462  
Db 214 aggttcggggagctccgcggccggccggccactcc-taccc-gcctacaggccg 272  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Qy 463 CCCCTCCGGGGACCTGCGGCGGGTCGCCTGCCAGCCCCCGAACCCGCGCTGGAGGCCG 522  
Db 273 cggccggctgggg--tcttagga--ctccg-ctgccggcgaaagagactccgccttgt 327  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Qy 523 CGGT CGGCCGGGCTCTCCGGAGGCACCCACTGCCACCGCGAAGAGTGGGC-TCTGT 581  
Db 328 cagccgcggg-cgcgcggggctggggccaggc-cggcgag-c--gccgcgaggacagg 382  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 582 CAGCCGCGGGTCTCTGGGGGCGAGGGCGAGGTTCAGGCCCTTCAGGCCGCAGGAAGAGG 641  
Db 383 aatggaactggtccccgtgttcggtgtc-ttacctgagctgtggaaagtgcacccggAAC 441  
||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 642 AACGGAGCGAGTCCCCGGCGCGCGCGATTCCCTGAGCTGTGGGACGTGCACCCAGGAC 701  
Db 442 tcggttctcaca 453  
||| ||| | | | |  
Qy 702 TCGGCTCACACA 713

RESULT 7  
LOCUS I28278 215 bp DNA PAT 30-OCT-1996  
DEFINITION Sequence 5 from patent US 5569830.  
ACCESSION I28278  
NID g1819054  
KEYWORDS .  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 215)  
AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.  
TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease  
JOURNAL Patent: US 5569830-A 5 29-OCT-1996;  
FEATURES Location/Qualifiers  
source 1..215  
/organism="unknown"  
BASE COUNT 15 a 8 c 25 g 26 t 141 others  
ORIGIN

```

Query Match           3.9%; Score 38; DB 71; Length 215;
Best Local Similarity 13.6%; Pred. No. 2.28e-10;
Matches 24; Conservative 79; Mismatches 71; Indels 3; Gaps 3;

Db      6 mssssvvsvrtascnkakkdgttsswttcdccnrwtgvcdtdttrvnnndsghnkyssan 65
       :: :::::: | : : : | : | : | : | || : : | : : : : | :::::
Qy    465 CCTCCCGGGGACCTGC GGCGGGT CGCCTGCCAGCCCCGAACCCGCTGGAGGCCGCG 524

Db      66 ynyggnnvgaakthyythtnv-sgadsktvtdsynasgtssnngtdgnrsgadsygssk 124
       :|| : | : ::|: : :| : : : | : : | : : | : :| : ::| : :
Qy    525 GTCGGCCCAGGGCTTCTCCGGAGGCACCCACTGCCACC-GCGAAGAGTTGGGCTGTCA 583

Db      125 tamtsrnrtgk-tannavdsrnmgdasvgskntkkhaknsadgkvgsknngdrnnr 180
       : : | : | : | : : | : : | : : | : : | : : | : : | : :
Qy    584 GCGCGGGTCTCGGGGGCGAGGGCGAGGTTCAGGCCCTTCAGGCCGCAGGAAGAG 640

```

RESULT 8  
 LOCUS I28278 215 bp DNA PAT 30-OCT-1996  
 DEFINITION Sequence 5 from patent US 5569830.  
 ACCESSION I28278  
 NID g1819054  
 KEYWORDS .  
 SOURCE Unknown.  
 ORGANISM Unknown.  
     Unclassified.  
 REFERENCE 1 (bases 1 to 215)  
 AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.  
 TITLE Plant inhibitors of fungal polygalacturonases and their use to  
     control fungal disease  
 JOURNAL Patent: US 5569830-A 5 29-OCT-1996;  
 FEATURES Location/Qualifiers  
     source 1..215  
           /organism="unknown"  
 BASE COUNT 15 a 8 c 25 g 26 t 141 others

ORIGIN

Query Match 3.6%; Score 35; DB 71; Length 215;  
Best Local Similarity 16.5%; Pred. No. 3.99e-08;  
Matches 34; Conservative 77; Mismatches 93; Indels 2; Gaps 2;

Db 10 svvsrtascnckdgntsswttccnrtwgvcldttrvnnndsghnkyssanyng 69  
:::|:::| | :: :| | :| :| :| :| :| :|  
Cp 866 GGCGACTTGGAGGTGCCCTCACGTCCTGCCAATTGCAGCACACTGGCCCAGTCAG 807

Db 70 gnnvgaakthythtnvsgadsktvtdsynasgtssssngtgdgnrsgadsgyssktamts 129  
:| :| :| :| :| :| :| :| :| :| :| :| :|  
Cp 806 TCAGGTTGGGGTTCAAAAGCCCCACTGCCGGCGAGGGGTGACGGATGCGCACGATCG 747

Db 130 rnrtgktannavdsrn-m-gdasvg sdkntkhaknsadgkvgsknngdrnnrygtgtsn 188  
:| :| :| :| :| :| :| :| :| :| :| :| :|  
Cp 746 GCGTCCCCCAACAGGAAAGCGAACTGCAT-GTGTGAGCCGAGTCCTGGGTGCACG 688

Db 189 vsnnccgggnkrdvssyannkccgssc 214  
:| :| :| :| :|  
Cp 687 TCCCCACAGCTCAGGGAATCGCGCCGC 662

RESULT 9

LOCUS I31769 30 bp DNA PAT 20-DEC-1996  
DEFINITION Sequence 22 from patent US 5583016.  
ACCESSION I31769  
NID g1822560  
KEYWORDS .  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Villeponteau,B., Feng,J., Funk,W. and Andrews,W.H.  
TITLE Mammalian telomerase  
JOURNAL Patent: US 5583016-A 22 10-DEC-1996;  
FEATURES Location/Qualifiers  
source 1..30  
/organism="unknown"  
BASE COUNT 13 a 6 c 11 g 0 t  
ORIGIN

Query Match 3.1%; Score 30; DB 72; Length 30;  
Best Local Similarity 100.0%; Pred. No. 1.45e-04;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gagaaaaacagcgcgcggggagcaaaagca 30  
|||||||||||||||||||||||||||||||  
Cp 371 GAGAAAAACAGCGCGCGGGGAGCAAAAGCA 342

RESULT 10

LOCUS I31770 26 bp DNA PAT 20-DEC-1996  
DEFINITION Sequence 23 from patent US 5583016.  
ACCESSION I31770  
NID g1822561  
KEYWORDS .  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Villeponteau,B., Feng,J., Funk,W. and Andrews,W.H.  
TITLE Mammalian telomerase  
JOURNAL Patent: US 5583016-A 23 10-DEC-1996;  
FEATURES Location/Qualifiers  
source 1..26

```

        /organism="unknown"
BASE COUNT      7 a      3 c      9 g      7 t
ORIGIN

Query Match      2.7%;  Score 26;  DB 72;  Length 26;
Best Local Similarity 100.0%;  Pred. No. 6.35e-02;
Matches 26;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Db      1 gtttgctctagaatgaacgggtgaaag 26
       ||||||| ||||||| ||||||| |||||
Cp      435 GTTTGCTCTAGAATGAACGGGTGAAAG 410

RESULT 11
ID HS366D1 standard; DNA; HTG; 54404 BP.
AC Z97986;
NI e1056722
DT 22-JUL-1997 (Rel. 52, Created)
DT 22-JUL-1997 (Rel. 52, Last updated, Version 1)
DE Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 366D1
KW HTG; HTGS_PHASE1.
OS Homo sapiens (human)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
OC Homo.
RN [1]
RP 1-54404
RA Barlow K. ;
RT ;
RL Submitted (20-JUL-1997) to the EMBL/GenBank/DDBJ databases.
RL Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA,
RL UK. E-mail enquires: humquery@sanger.ac.uk Clone requests:
RL clonerequest@sanger.ac.uk
CC IMPORTANT: This sequence is unfinished and does not necessarily
CC represent the correct sequence. Work on the sequence is in progress
CC and
CC the release of this data is based on the understanding that the
CC sequence may change as work continues. The sequence may be
CC contaminated
CC with foreign sequence from E.coli, yeast, vector, phage etc.
CC Order of segments is not known; 800 n's separate segments.
CC Unfinished sequence: c366D1 Contig_ID: 00051 Length: 6814 bp
CC Unfinished sequence: c366D1 Contig_ID: 00729 Length: 16382 bp
CC Unfinished sequence: c366D1 Contig_ID: 00410 Length: 7092 bp
CC Unfinished sequence: c366D1 Contig_ID: 01222 Length: 21716 bp
FH Key          Location/Qualifiers
FH
FT source      1..54404
FT           /organism="Homo sapiens"
FT           /clone="366D1"
FT           /chromosome="16"
SQ Sequence 54404 BP; 9401 A; 15916 C; 16477 G; 10164 T; 2446 other;

Query Match      2.7%;  Score 26;  DB 5;  Length 54404;
Best Local Similarity 83.0%;  Pred. No. 6.35e-02;
Matches 39;  Conservative 0;  Mismatches 7;  Indels 1;  Gaps 1;

Db 50565 cgggggtggggcggggtgggcgggcgaccggccgc-ggtccccgcgagg 50610
       |||||  || | ||| ||| ||| ||| ||| ||| ||| ||| |
Cp 511 CGGGGTTCGGGGGCTGGGCAGGCGACCCGCCGCAGGTCCCCGGGAGG 465

RESULT 12
LOCUS HS366D1 54404 bp DNA HTG 22-JUL-1997
DEFINITION Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 366D1;
HTGS phase 1.

```

ACCESSION Z97986  
 NID g2274954  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;  
 Homo.  
 REFERENCE 1 (bases 1 to 54404)  
 AUTHORS Barlow, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-1997) Wellcome Trust Genome Campus, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquires:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 COMMENT IMPORTANT: This sequence is unfinished and does not necessarily  
 represent the correct sequence. Work on the sequence is in progress  
 and  
 the release of this data is based on the understanding that the  
 sequence may change as work continues. The sequence may be  
 contaminated  
 with foreign sequence from E.coli, yeast, vector, phage etc. Order  
 of segments is not known; 800 n's separate segments. Unfinished  
 sequence: c366D1 Contig\_ID: 00051 Length: 6814 bp Unfinished  
 sequence: c366D1 Contig\_ID: 00729 Length: 16382 bp Unfinished  
 sequence: c366D1 Contig\_ID: 00410 Length: 7092 bp Unfinished  
 sequence: c366D1 Contig\_ID: 01222 Length: 21716 bp.  
 \*\*\*  
 \*\*\* WARNING: Phase 1 High Throughput Genome Sequence \*\*\*  
 \*\*\*  
 \* This sequence is unfinished. When sequencing is complete,  
 \* the sequence data presented in this record will be replaced  
 \* by a single finished sequence with the same accession number.  
 FEATURES Location/Qualifiers  
 source 1..54404  
     /organism="Homo sapiens"  
     /clone="366D1"  
     /chromosome="16"  
 BASE COUNT 9401 a 15916 c 16477 g 10164 t 2446 others  
 ORIGIN  
 Query Match 2.7%; Score 26; DB 128; Length 54404;  
 Best Local Similarity 83.0%; Pred. No. 6.35e-02;  
 Matches 39; Conservative 0; Mismatches 7; Indels 1; Gaps 1;  
 Db 50565 cgggggtggggcggggtggcgccgcgaccggccgc-ggtcccccgcgagg 50610  
     ||||||| ||| ||| ||||||| ||||||| ||||| ||||||||| |||||  
 Cp 511 CGGGGGTTCGGGGGCTGGGCAGGCGACCCGCCGCAGGTCCCCGGGAGG 465  
 RESULT 13  
 LOCUS E04076 565 bp RNA PAT 26-NOV-1996  
 DEFINITION gDNA encoding envelope region of type C hepatitis virus.  
 ACCESSION E04076  
 NID g2172286  
 KEYWORDS JP 1992349885-A/1.  
 SOURCE Hepatitis C virus.  
 ORGANISM Hepatitis C virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Hepatitis C-like viruses.  
 REFERENCE 1 (bases 1 to 565)  
 AUTHORS Tsutae,M., Kazuaki,C., Hiromitsu,K. and Yataro,I. .  
 TITLE NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS  
 AND METHOD FOR DETECTING THE SAME  
 JOURNAL Patent: JP 1992349885-A 1 04-DEC-1992;  
 TEIJIN LTD  
 COMMENT OS Hepatitis C virus

PN JP 1992349885-A/1  
 PD 04-DEC-1992  
 PF 29-MAY-1991 JP 1991152169  
 PI MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUMADA HIROMITSU, PI  
 ICHIKAWA YATARO  
 PC C12N15/10,C12Q1/68,C12Q1/70//C12N15/11;  
 CC strandedness: Single;  
 CC topology: Linear;.  
**FEATURES** Location/Qualifiers  
 source 1..565  
     /organism="Hepatitis C virus"  
**BASE COUNT** 60 a 93 c 107 g 85 t 220 others  
**ORIGIN**  
 Query Match 2.5%; Score 25; DB 68; Length 565;  
 Best Local Similarity 20.3%; Pred. No. 2.68e-01;  
 Matches 26; Conservative 49; Mismatches 53; Indels 0; Gaps 0;  
 Db 199 rargyrrmrggayryvatcmtgcwydyccbgggtgyrybccytgygtcgsarrryrry 258  
   :|:|:::: :::: || : :||:|||||::: |: : |:::: :::  
 Cp 726 AAAGCGAACTGCATGTGTGAGCCGAGTCCTGGGTGCACGTCCCACAGCTCAGGGAATCGC 667  
 Db 259 nnytcbmgntgytgggyrgcgctbacycacgstygcvrcyagrayvbyavyvtyccc 318  
   :|::| | :: | :| :| | | : | :: | :: | ::| :| |  
 Cp 666 GCCGCCGCAGGGACTCGCTCCGTTCTCCTGCAGGCTGAAAGGCCTGAACCTCGCC 607  
 Db 319 rcbrygv 326  
   ::: :::  
 Cp 606 CTCGCC 599  
  
**RESULT** 14  
**LOCUS** HUMA2PIG1 1658 bp DNA PRI 30-OCT-1994  
**DEFINITION** Human alpha-2-plasmin inhibitor gene, exon 1.  
**ACCESSION** M20781 J03830  
**NID** g177879  
**KEYWORDS** alpha-2-plasmin inhibitor; plasmin inhibitor; serine protease inhibitor; serpin.  
**SEGMENT** 1 of 6  
**SOURCE** Human placenta DNA, clones lambda-PI[1,2,6].  
**ORGANISM** Homo sapiens  
   Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
   Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 1658)  
**AUTHORS** Hirose,S., Nakamura,Y., Miura,O., Sumi,Y. and Aoki,N.  
**TITLE** Organization of the human alpha 2-plasmin inhibitor gene [published erratum appears in Proc Natl Acad Sci U S A 1989 Mar;86(5):1612-1]  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 85 (18), 6836-6840 (1988)  
**MEDLINE** 88320531  
**COMMENT** Draft entry and computer-readable sequence for [1] kindly submitted by N.Aoki, 28-SEP-1988.  
**FEATURES** Location/Qualifiers  
 source 1..1658  
     /organism="Homo sapiens"  
     /tissue\_type="placenta"  
     /map="18p11.1-q11.2"  
 TATA\_signal 1070..1075  
     /gene="PLI"  
     /note="G00-120-301"  
 exon 1087..1104  
     /gene="PLI"  
     /note="G00-120-301"  
     /number=1  
 intron 1105..>1658  
     /gene="PLI"  
     /note="G00-120-301"

/number=1

BASE COUNT 282 a 495 c 533 g 348 t  
 ORIGIN 233 bp upstream of BamHI site.

Query Match 2.5%; Score 25; DB 97; Length 1658;  
 Best Local Similarity 69.8%; Pred. No. 2.68e-01;  
 Matches 44; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

Db    633 gcaggactctggcctgtgagggtgggttctggctttcatgccccctgatgagggtcag 692
      ||| ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    587 GCGGGTCTCTCGGGGGCGAGGGCGAGGTTCAGGCCCTTCAGGCCGCAGGAAGAGGAACGG 646
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Db    693 agc 695
      |||
Qy    647 AGC 649
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

```

RESULT 15

Locus PSESDSAB11 3510 bp DNA BCT 04-JUN-1992

DEFINITION Pseudomonas sp. (strain ATCC 19151) sdsA gene and 11 kd protein, complete cds's; sdsB gene, partial cds.

ACCESSION M86744

NID g151550

KEYWORDS 11 kd protein; alkyl sulfatase; positive transcriptional control factor.

SOURCE Pseudomonas sp. DNA.

ORGANISM Pseudomonas sp.

Eubacteria; Proteobacteria.

REFERENCE 1 (bases 1 to 3510)

AUTHORS Davison,J., Brunel,F., Phanopoulos,A., Prozzi,D. and Terpstra,P.

TITLE Cloning and sequencing of Pseudomonas genes determining sodium dodecyl sulfate biodegradation

JOURNAL Gene 114, 19-24 (1992)

MEDLINE 92267380

FEATURES Location/Qualifiers

source	1..3510
	/organism="Pseudomonas sp."
gene	complement(57..977)
	/gene="sdsB"
CDS	complement(57..977)
	/gene="sdsB"
	/codon_start=1
	/function="activates sdsA transcription"
	/evidence=experimental
	/product="positive transcriptional control factor"
	/db_xref="PID:g151551"
	/transl_table=11
	/translation="MNDLRQLRHVALAEHGHFARAAEAVNLSPALRSIQAENGGL
	GCRLLDRGPRQVSLTAHGRRLVLEHARRLLDGDRALRSAVSQDNLGSSELRLGAGPYP
	GARLVPRALGRFAGAHPGVRVQLAIDTWYSLHQQLDDALELFVADRELRRDPQLEV
	TPLRSWPGVIFCRPGHPLLGRRLTAADLAAYPLAGTQVPAEVQAQALGQLAESGQP
	LGIECDNFMALKAVALVAESDVLSMAPLDVVAAEIEAGRLALLELAPGLLSQRSAYGLVS
	RAGRTLSPAAEAMRGLILDEDARTPPASAR"
CDS	1276..2862
	/gene="sdsA"
	/codon_start=1
	/function="degrades sulfate esters of long-chain primary alcohols (SDS)"
	/evidence=experimental
	/product="alkyl sulfatase"
	/db_xref="PID:g151552"
	/transl_table=11
	/translation="MIEAPEGLIIVDTGESVDQSRKVLAEFRKISDKPIKAIVYTHFH
	PDHINGVKAFVSEEQVKSGEVRIYAQETLLDNVTQGSLVGPILTMRSGYSGVALSD
	EDKRDMNAGLGPLAHEGASTFIAPTDTFRDSLDTTIAGLKQFLHVSEAPDEIVLYL
	PDNRVLISAEVTQGPTLPNVHTLRGTFRDPPVWWVASLDKLRAFQADVMVPLHGQPVS

GREKVEEVLRMTRDAIAYIHDTQTVRWMNKGLTPDELVEVKLPPHLAGYTPYLREYYG  
 TVKHSVRQIYQGYLGWFQGDPVLDPIPPEAKARRLIALMGGRKVLMAAGDAYLKGD  
 WQWAAEELSGYAIRVDHDDKLARDIKARSFRLGYASMNINWRNWYLM SAMELEGKLEG  
 DVALEMSRRVRAAFLSPDMLKNLPARIFLQNWWTRIDPEKSGDVELALGFAFPDIDEA  
 WTLEVRRGVAQLKSGIDPAVPLRLTLKRYLDTVISGENSLKGALLGDVKVDGNLLD  
 IKTFLGCDFDAPIALTVR"  
 gene 1276..2862  
 /gene="sdsA"  
 CDS 2955..3200  
 /note="putative"  
 /codon\_start=1  
 /function="unknown"  
 /product="11 kd protein"  
 /db\_xref="PID:g151553"  
 /transl\_table=11  
 /translation="MHLPNFLKGALAMFWLVALLNVIYPFGETLHRPLLLASALVLLA  
 HVGEAALFNRKLRARPSPWLERIQVLLFGFLHLRGLR"  
 BASE COUNT 579 a 1252 c 1150 g 529 t  
 ORIGIN  
 Query Match 2.5%; Score 25; DB 32; Length 3510;  
 Best Local Similarity 69.7%; Pred. No. 2.68e-01;  
 Matches 53; Conservative 0; Mismatches 22; Indels 1; Gaps 1;  
 Db 631 ccggcgaaggcgccccagcgccggggca-ccaggcgccgcgggttagggccggacc 689  
 ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Cp 545 CCGGAGAACCCCCGGGCCGACCGCGGCCCTCCAGGCGGGTTCGGGGCTGGGCAGGCGAC 486  
 Db 690 caggcgcagttcgccg 705  
 ||| ||||| ||| |||  
 Cp 485 CCGCCGCAGGTCCCCG 470

Search completed: Thu Feb 5 05:30:23 1998  
 Job time : 1615 secs.